Data Manipulation in the Undergraduate Laboratory: What are we teaching?

Elizabeth W. Davidson, Department of Biology, Arizona State University, USA

Heather E. Cate, College of Liberal Arts and Sciences, Arizona State University, USA

Cecil M. Lewis, Jr., Department of Anthropology, University of New Mexico, USA

Melanie Hunter, Department of Biology, Arizona State University, USA

Keywords: Biology, Cheating, Chemistry, Motivation, Undergraduate research

Arizona State University (ASU) offers a senior-level course entitled "Professional Values in Science" that addresses a number of topics concerning ethical conduct of research as well as ethical concerns at the intersection of science and society. The course demands active participation by the students. Several years ago, on his own initiative a student in the class developed a questionnaire that explored data manipulation. As most of the students were undergraduates, the questionnaire focused upon manipulation of data in undergraduate science laboratories. We were startled to discover that over 60% of the students openly admitted to manipulation of data in undergraduate laboratories. These results led to the development of a more elaborate survey that has been administered to 7 undergraduate Biology and Chemistry courses, enrolling a total of over 700 students. The courses include both major and nonmajor subjects, at both introductory and upper division level. Arizona State University has approximately 50,000 students, including (in academic year 2000) ca. 1000 majors in Biology and 250 majors in Chemistry. In the fall semester, 2000, 3137 undergraduates are enrolled in Biology courses, while 3355 undergraduates are enrolled in Chemistry courses. Laboratories are therefore limited in available time, are generally supervised and graded by graduate teaching assistants, and many, but not all, of these courses rely upon traditional laboratory exercises.

Methods:

The survey and instructions to students are presented in at the end of the paper. Students were advised by the person administering the survey (who was not their course professor or teaching assistant) that the results would be held anonymous and would not affect their grade. The courses included Chemistry 115: Introductory, non-majors; Chemistry 335: Organic, non-majors; Biology 201: Anatomy and Physiology, non-majors; Biology 100: Introductory, non-majors; Biology 182: Introductory, majors; Biology 193: Introductory, majors, critical thinking focus; Biology 385: Invertebrate Zoology, majors. Seven hundred and two students participated. Institutional Human Subjects committee approval was obtained. Data were analysed by Spearman correlation.

Corresponding author: Elizabeth W. Davidson, Department of Biology, Arizona State University, Box 871501, Tempe, AZ 85287-1501, 480-965-7560 (voice), 480-965-2519 (fax), e.davidson@asu.edu.

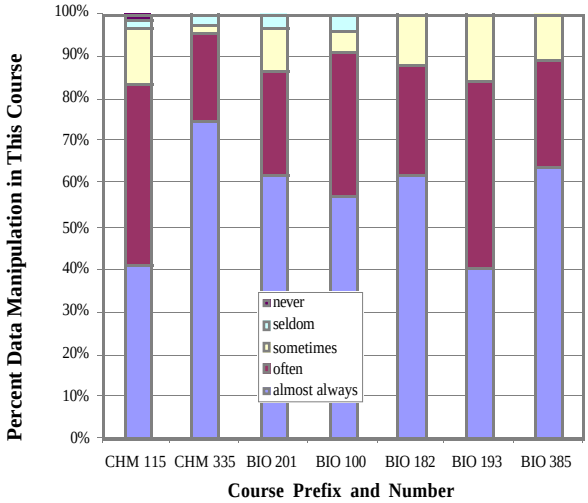

Figure 1. Results of survey, Question 5, "Have you ever manipulated data or made up data in this course?" CHM 115: Introductory, non-majors, N=86; CHM 335: Organic, non-majors, N=44; BIO 201: Anatomy and Physiology, non-majors, N=29; BIO 100: Introductory, non-majors, N=200; BIO 182: Introductory, majors, N=40; BIO 193: Introductory, majors, critical thinking focus, N=57; BIO 385: Invertebrate Zoology, majors, N=64. N= total number of responses to the specific question.

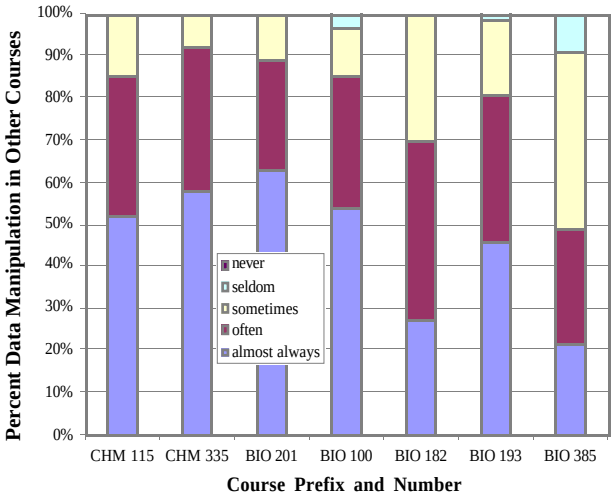

Figure 2. Results of survey, Question 10, "Have you ever manipulated or made up data in any other *science course?" CHM 115, N=87; CHM 335, N=52; BIO 201, N=27; BIO 100, N=81; BIO 182, N=40; BIO 193, N=57; BIO 385, N=66. N= total number of responses to the specific question.*

Results

The key question in this survey was Question 5, "Have you ever manipulated data in this course?" As shown in Figure 1, between 40.4 and 75% of students in the surveyed course admitted to manipulating data "almost always," and another 20-43.9% admitted to such manipulation "often." Students reporting data manipulation "seldom" represented less than 5% of those surveyed, and only one student out of over 500 who responded to this question replied "never." Using correlation analysis, we learned that admission of manipulation in the course surveyed was strongly correlated to admission of manipulation in other courses (Spearman Correlation Sig. (2 tailed) 0.355, significant at 0.01 level) (Figure 2).

We asked whether data manipulation was related to the level (i.e. introductory vs. advanced) of the course, and whether the course was designed for majors or non-majors. No significant difference was found between data manipulation in Introductory Biology BIO 100 (non-majors) and BIO 182 (majors) or between these lower division courses and an upper division course, BIO 385 (Invertebrate Zoology, majors). We compared responses from BIO 182, a traditional introductory course, to BIO 193, an introductory majors course with emphasis on critical thinking. The smallest percentage of students reporting data manipulation "almost always" was in BIO 193, however a large proportion of the remainder reported manipulation "often" (Figure 1). Within the two non-majors chemistry courses surveyed, less data manipulation was found in CHM 115 (Introductory) than in CHM 335 (Organic), and indeed the highest overall reported manipulation (90.5% "almost always" or "often") was reported in Organic Chemistry. Conversations with students in the Professional Values in Science class and elsewhere confirmed that many have

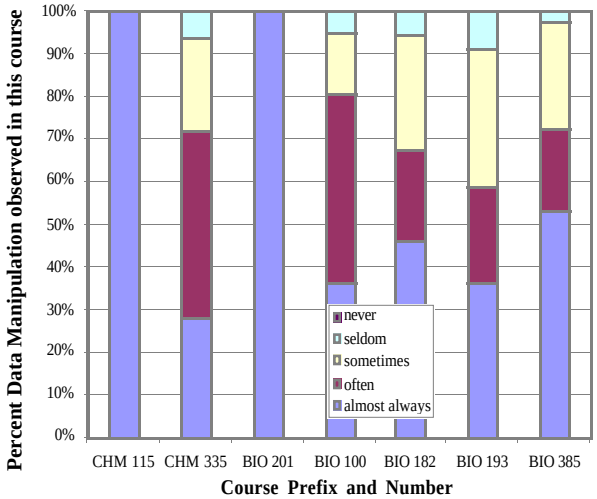

Figure 3. Results of survey, Question 7, "Have you ever observed anyone else *manipulate or make up data in* this course*?" CHM 115, N=91; CHM 335, N=67; BIO 201, N=28; BIO 100, N=237; BIO 182, N=40; BIO 193, N=56; BIO 385, N=66. N= total number of responses to the specific question.*

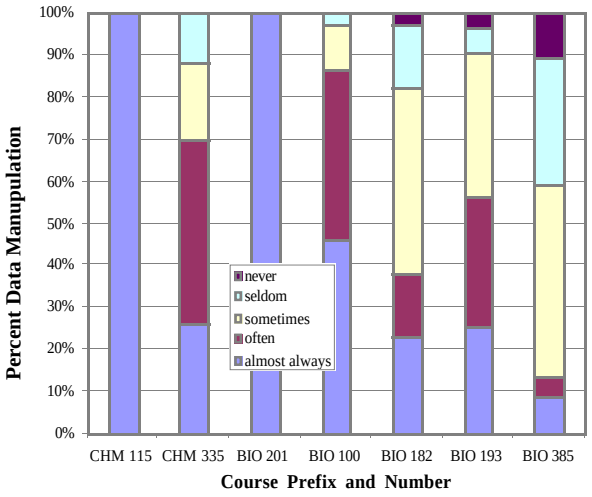

Figure 4. Results of survey, Question 14, "Have you ever observed anyone manipulate or make up data in any *science course?" CHM 115, N=94; CHM 335, N=70; BIO 201, N=30; BIO 100, N=96; BIO 182, N=39; BIO 193, N=55; BIO 385, N=66. N= total number of responses to the specific question.*

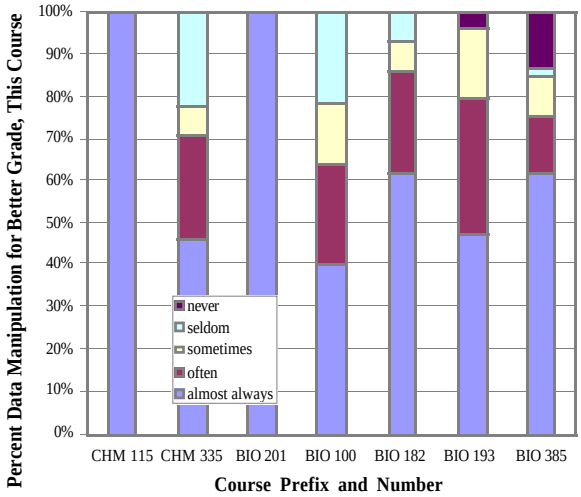

Figure 5. Results of survey, Question 6, "If you have ever manipulated data or made up data, was it motivated by the thought of a better grade?" CHM 115, N=69; CHM 335, N=41; BIO 201, N=17; BIO 100, N=246; BIO 182, N=31; BIO 193, N=55; BIO 385, N=53. N= total number of responses to the specific question.

manipulated data for Chemistry laboratory reports, particularly in Organic. Little difference in data manipulation (Question 5) was found when analyzed by academic year or by gender.

Two other key questions were 7 and 14, which asked whether the student had observed others manipulating data. The results from these questions were less consistent than responses about the students own data manipulation. Two courses (CHM 115 and BIO 201) received an "almost always" response rate of 100%, whereas in other courses a much smaller proportion of students responded "almost always" (Figures 3, 4).

We investigated motivation for data manipulation with questions 6 and 11, which asked whether the students manipulated data in order to get a better grade. Up to 100% of students in some courses replied that manipulation was almost always performed to obtain a better grade (Spearman Correlation Sig. (2-tailed) 0.265, significant at 0.01 level) (Figure 5; data from Question 11 not shown). When asked whether this was because they felt that their grade depended heavily upon the experimental results (Questions 8 and 15), less than half of students felt that their grade in the current course depended on experimental results "almost always", and from 3.0 to 13.6% of the students replied to Question 8 that their grade "seldom" depended on results (Figure 6, data from Question 15 not shown; Spearman (Figure 7, data from Question 16 not shown; Spearman correlation 0.368, significant at 0.01 level). Finally we surveyed student preferences for type of laboratory experiments (Question 17). In all seven courses combined, only 1.7% of students preferred lab experiments which place more emphasis on results, whereas 53.5% preferred more emphasis to be placed upon processes, and 44.7% preferred a balanced combination of both techniques (N=503).

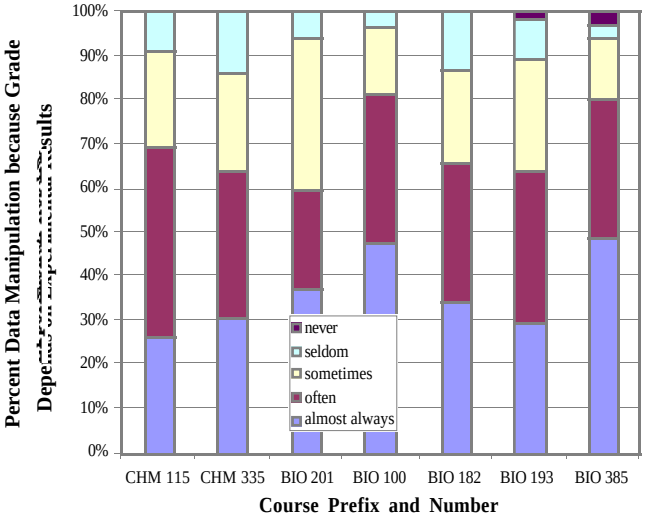

Figure 6. Results of survey, Question 8, "How often have you felt as though your grade in this course depended heavily on your experimental results?" CHM 115, N=102; CHM 335, N=66; BIO 201, N=35; BIO 100, N=218; BIO 182, N=40; BIO 193, N=58; BIO 385, N=66. N= total number of responses to the specific question.

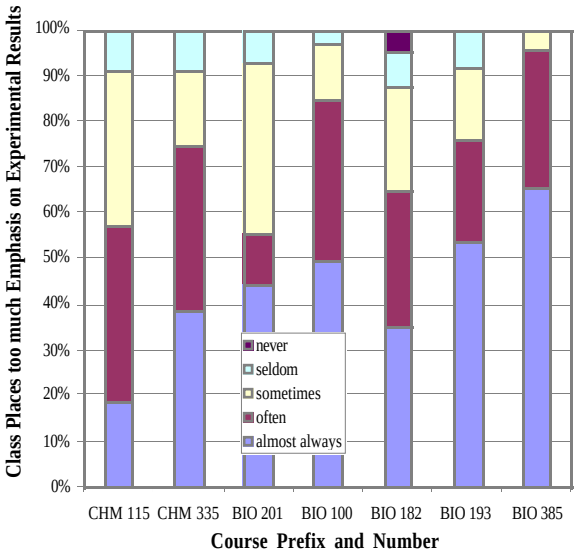

Figure 7. Results of survey, Question 9, "Do you believe this course places too much emphasis on experimental results rather than on the processes used to get the results?" CHM 115, N=98; CHM 335, N=67; BIO 201, N=27; BIO 100, N=194; BIO 182, N=40; BIO 193, N=58; BIO 385, N=66. N= total number of responses to the specific question.

Discussion:

Some precautions should be taken in interpreting these findings. First, the survey was limited to only 7 courses at a single University, which in each case were surveyed only once. We intend to survey additional courses at ASU, and hope to include at least one other large university and a small liberal arts college in our analysis. Second, the survey relies on self reporting. Some of the students did not choose to answer all questions in the survey. The total number responding to each question in each course is presented in the figure caption. Approximately 25% of the students chose not to answer Question 5, for example. Third, the construction of the questions did not permit us to investigate motivations other than that the student felt his/her grade depended upon the experimental results (Questions 8, 9, 15 - 17). Finally, even though students were given a clear definition of "data manipulation" at the beginning of the survey, it is possible that some may not have clearly understood the definition of "data manipulation."

With the above caveats in mind, our results show a very strong tendency among undergraduate science students to manipulate or make up data when writing laboratory reports. As high as these percentages are, they are similar to results observed in surveys of cheating on tests, which Cizek has described as "remarkably and uniformly high" (1). In surveys taken from 1970 to present, from 42% to over 90% of students reported cheating on tests by self or others (reviewed by Cizek, (1)). Out of 6000 students in 31 universities surveyed by Meade, 67% of science majors reported cheating on tests (2). Most surveys of college test cheating ask only whether the student has ever cheated. Our survey expands this question to evaluate how frequently data manipulation in laboratory reports occurs, allowing us to differentiate between occasional events and habitual cheating. Although there are many studies of cheating on college tests, to our knowledge our study is unique in examining manipulation of laboratory data by undergraduates.

Data manipulation apparently does not diminish as the students progress to upper division courses or from non-major to major courses. Commitment to a major subject, presumably because the student intends to continue in this area of science for all or part of his/her professional career, apparently does not diminish this practice.

These results raise some important questions, which include: How can this data manipulation be reduced or eliminated? What are the implications of data manipulation in the undergraduate laboratory to the future careers of these students? In other words, when do the students stop manipulating data? In graduate, professional or medical school? When they begin doing "real" research? When the research is published?

In response to the first of these questions, the faculty and the system itself must take significant responsibility. Faculty must recognize that this data manipulation occurs, and not turn a blind eye to this practice. We must examine the reason why we require laboratory reports in the first place, and whether there is another method of assessing whether the student has learned the necessary laboratory skills. Numerous laboratory manuals are structured to provide "cook book" procedures in which students are expected to verify known biological, chemical, or physical laws (3). However, these verification exercises give students a false notion of the deductive investigative process. They begin their training with the preconceived notion that a "right" answer exists and should be found. They are therefore willing to adjust their laboratory results for fear that the "wrong" answer would affect their grade (4).

We must change the common perception among undergraduate students that their grade often depends upon producing the "right" answer (Figure 6). This change will involve not only the laboratory experimental design, but also the training of graduate teaching assistants and elimination of grading based on achieving a preconceived result. Although students must still be evaluated on whether they are using laboratory instruments accurately, we must consider whether a given laboratory can be designed for training in the hypothetical-deductive process in addition to the specific laboratory technique (4, 5).

Unfortunately, the number of students enrolled in science laboratory courses at large universities in many ways promotes cook-book laboratory exercises. The limited time allowed for experiments, inability to repeat an experiment, and disinterest of many teaching assistants in spending adequate time to grade reports all contribute to the perception on the part of students that making up data is more profitable than writing up what really happened.

Faculty must rethink the reasons for requiring laboratory reports. If the reasons include determining whether the student was present, completed the tasks, understood the experiment, and learned the techniques, then the results presented here suggest that these goals are not being accomplished by the current mechanism of laboratory reports graded based upon achieving the "right" answer. Other mechanisms for discovering whether students have learned the important aspects of the exercise may include laboratory-based questions on exams, and building later experiments upon earlier laboratory exercises. Instructors must be willing to address this problem bluntly with the students and teaching assistants.

At ASU some laboratories have been redesigned to emphasize the inquiry approach to laboratories in Biology and Chemistry. Students generate alternative hypotheses and design experiments themselves, and concepts are introduced after, not before, results are obtained. Teaching assistants are trained to lead students into open-ended and thought-provoking questions (4, 5). In spite of these efforts, our data suggest that data manipulation occurs in these laboratories as well. As students commonly state, "everybody does it." The students themselves overwhelmingly prefer laboratory exercises which emphasize processes or a balance between process and results.

The second concern, whether undergraduates continue data manipulation as professional scientists, has even greater implications. In the frequently-cited study by Swazey et al., 72% of graduate students and 59% of faculty reported to have observed or had direct evidence of some form of scientific misconduct (6). Data falsification, however, was reported by a much smaller proportion of respondents, ranging from 2% to 20%. Apparently, then, data manipulation does decrease when the student becomes a "professional" and becomes dedicated to the science.

Over the last 5 years approximately 400 undergraduates at ASU have been engaged in research programs funded by the Howard Hughes Medical Foundation, the National Institutes of Health and the National Science Foundation, in addition to individual faculty grants. Conversations with these students reveal that once the research becomes "their own" and important to them personally, they have far less motivation to manipulate data, particularly if they have a respectful relationship with the faculty mentor. Hands-on undergraduate research experience may therefore be important in molding the ethical practices of students who will go on to become professional scientists.

When we emphasize the importance of getting the "right" answer, we are teaching undergraduates that their hypothesis must be supported. In truth, the function of an experiment should be to *allow for a fair test* of the hypothesis. We recognize that there exists temptation for graduate students and professional scientists to manipulate data in order to finish research before a deadline, to obtain the next grant, or to have an outstanding publication record. We must take serious responsibility that we do not teach data manipulation techniques at the undergraduate level that will continue to be used in later professional careers.

Acknowledgements

We are grateful to Dr. Anton Lawson for helpful suggestions.

Bibliography

1. Cizek, GJ. Cheating on tests: How to do it, detect it, and prevent it. L Erlbaum Assoc, Mahwah, NJ. 1999.
2. Meade J. Cheating: Is academic dishonesty par for the course? Prism 1992;1: 30-32.
3. Germann PJ, Haskins S, Auls S. Analysis of nine high

school biology laboratory manuals: promoting scientific inquiry. J Res Science Teaching 1996; 33: 475-499.
4. Lawson AE, Lewis CM Jr., Birk JP. Why do students "cook" data? J College Science Teaching 2000; 29(3): 191-198.
5. Lawson AE, Rissing SW, Faeth SH. An inquiry approach to nonmajors biology. J. College Science Teaching 1990;19(6): 340-346.
6. Swazey JP, Anderson MS, Lewis KS. Ethical problems in academic research. Amer Scientist 1993; 81: 342-553.

Data Manipulation Survey

Instructions to students:
Space shuttles blow up, bridges fall, and planes crash and not all are due to natural disasters. An undergraduate student at ASU has been conducting a research project for the last year and a half. During his undergraduate career, he found that in some laboratory settings, there appears to be a great deal of pressure to get the "right" result rather than an emphasis on the scientific and experimental process. In one of his labs he found that 80% of the students manipulated data in some way during the semester. He became concerned: where do students learn scientific ethics? Should we have faith that human morality will overcome pressures to manipulate data in the hopes of a better grade in our college career, or a publication in our professional career?

The purpose of this survey is to collect data on the extent to which undergraduates feel pressured to manipulate, change, or make up data acquired in the laboratory. For example, if you only have a 30% yield of a particular reaction, have you ever felt pressured to say you had more to get a better grade? Moreover, how did you respond to that pressure? Alternatively, has the lab concentrated on experimental process rather than actual results?

Data Manipulation**: To change or omit acquired data or to make up data without confession to those evaluating your performance.**

1. What is your TA's name?

2. What is you major and what year are you (freshman, sophomore, etc.)?

3. Are you:
   A. Female B. Male

4. How many science labs have you taken?
   A. 1 B. 2-5 C. 6 or more

5. Have you ever manipulated data or made up data in this course?
   A. Almost Always B. Often C. Sometimes D. Seldom E. Never
6. If you have ever manipulated data or made up data, was it motivated by the thought of a better grade?
   A. Almost Always B. Often C. Sometimes D. Seldom E. Never

7. Have you ever observed anyone else manipulate or make up data in this course?
   A. Almost Always B. Often C. Sometimes D. Seldom E. Never

8. How often have you felt as though your grade in this course depended heavily on your experimental results?
   A. Almost Always B. Often C. Sometimes D. Seldom E. Never

9. Do you believe this course places too much emphasis on experimental results rather than on the processes used to get the results?
   A. Almost Always B. Often C. Sometimes D. Seldom E. Never

10. Have you ever manipulated or made up data in any other science course?
   A. Almost Always B. Often C. Sometimes D. Seldom E. Never

11. If you have manipulated or made up data, was it motivated by the thought of a better grade?
A. Almost Always B. Often C. Sometimes D. Seldom E. Never

12. If you have manipulated or made up data, was (were) the course(s):
A. Lower Division (100-200 level) B. Upper Division (300 or 400 level) C. Both A & B

13. If you have manipulated or made up data, what department was (were) the course(s) in? (Please circle all that apply.)
A. Biology B. Physics C. Chemistry D. Zoology E. Botany F. Microbiology

14. Have you ever observed anyone manipulate or make up data in any science course?
A. Almost Always B. Often C. Sometimes D. Seldom E. Never

15. How often have you felt that your grade in a science course depended heavily on you experimental results?
A. Almost Always B. Often C. Sometimes D. Seldom E. Never

16. Do you believe that science courses place too much emphasis on experimental results rather than on the processes used to get those results?
A. Almost Always B. Often C. Sometimes D. Seldom E. Never

17. Would you like to see lab experiments:
A. Place more emphasis on results. B. Place more emphasis on processes.
C. Have a balanced combination of both.